DOCUMENT RESUME

ED 324 647 CS 010 266

AUTHOR	Stallman, Anne C.; Pearson, P. David
TITLE	Formal Measures of Early Literacy. Technical Report No. 511.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Sep 90
CONTRACT	G0087-C1001-90
NOTE	52p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Cognitive Development; Kindergarten; Primary Education; Reading Diagnosis; *Reading Readiness; *Reading Readiness Tests; Reading Research; Reading Skills; Reading Tests; *Test Validity
IDENTIFIERS	Concepts About Print Test; *Emergent Literacy; Gesell School Readiness Test; Metropolitan Readiness Tests

ABSTRACT

This report reviews and evaluates the tradition of assessing early literacy with formal tests. It begins with a historical account of the readiness movement, focusing on the development of formal measures of early literacy and tracing the development of the movement into the 1980s. Next, the paper analyzes in depth tests that are currently available. Based on this analysis, the paper concludes that commercially available tests measure tasks that few would label literacy and recommends that future tests be designed to reflect both the emergent literacy tradition and the belief that teachers are professionals who need the best information possible to make informed decisions about children. Eight tables of data and 14 figures are included, and 43 references are attached. (Author/SR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED324647

CENTER FOR THE STUDY OF READING

Technical Report No. 511

FORMAL MEASURES OF EARLY LITERACY

Anne C. Stallman
&
P. David Pearson
University of Illinois at Urbana-Champaign

September 1990

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS010266

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

EDITORIAL ADVISORY BOARD
1989-90

James Armstrong
Linda Asmussen
Gerald Arnold
Yahaya Bello
Diane Bottomley
Catherine Burnham
Candace Clark
Michelle Commeyras
John M. Consalvi
Christopher Currie
Irene-Anna Diakidoy
Barbara Hancin
Michael J. Jacobson
Jihn-Chang Jehng
Robert T. Jimenez
Bonnie M. Kerr
Paul W. Kerr
Juan Moran
Keisuke Ohtsuka
Kathy Meyer Reimer
Hua Shu
Anne Stallman
Marty Waggoner
Janelle Weinzierl
Pamela Winsor
Marsha Wise

MANAGING EDITOR
Fran Lehr

MANUSCRIPT PRODUCTION ASSISTANTS
Delores Plowman
Debra Gough

Abstract

In this report, we review and evaluate the tradition of assessing early literacy with formal tests. We begin with a historical account of the readiness movement, focusing on the development of formal measures of early literacy and tracing the development of the movement into the 1980s. Next, we analyze in depth tests that are currently available. Based on this analysis, we conclude that commercially available tests measure tasks that few would label literacy, and we recommend that future tests be designed to reflect both the emergent literacy tradition and the belief that teachers are professionals who need the best information possible to make informed decisions about children.

FORMAL MEASURES OF EARLY LITERACY

The purpose of this report is to describe and evaluate formal measures of early literacy in relation to changing perspectives on how young children learn to read and write. Our ultimate goal is to be able to make recommendations about how, if at all, such measures should be used by educators.

Our plan is to trace the historical relationship of literacy measures to the child development theories. The historical perspective will serve as a foundation for the major focus of the report--a detailed analysis of those measures currently available for educators to use in making decisions about young children's "readiness" for literacy instruction. That analysis will provide the evidence to support whatever conclusions we reach about their potential usefulness.

The History

While concern for the instruction of young children can be traced back to the early Greeks--Aristotle and Plato advocated teaching fables to young children in order to promote goodness and development of the soul (Mason, 1984), the concept of "reading readiness" is a relative newcomer to the field of education. The ancient Greeks were not concerned about prereading skills because learning to read was not considered to be a task difficult enough to merit a readiness stage (Venezky, 1975). In fact, Socrates described learning to read in this way.

> Just as in learning to read, I said, we were satisfied when we knew the letters of the alphabet, which were very few, in all their recurring sizes and combinations; not slighting them as unimportant whether they occupy a space large or small, but everywhere eager to make them out; and not thinking ourselves perfect in the art of reading until we recognized them wherever they are found (Plato, 1892, III, 402a).

This view of reading prevailed until the 1900s when Huey (1908) put forth the idea that there was more to learning to read than alphabet identification. Although Huey's ideas represented a shift in thinking about the skills involved in the reading process, little attention was given to *when* reading instruction should begin. Many educators, including Huey and John Dewey, thought that age 8 was optimal for beginning reading instruction because then the child might have a rich enough store of concepts to be able to understand the ideas in texts.

Beginning in the 1920s, as a result of the scientific measurement and testing movement, educators became concerned about the number of children who were unsuccessful in first grade, usually due to a lack of reading skills. There were many studies conducted during the 1920s and 1930s to determine the reasons for children's failure in first grade (e.g., Dickson, 1920; Holmes, 1927; Monroe, 1932; Smith, 1928). This type of research is typified by a study by Morphett and Washburne (1931), which recommended that teachers could reduce the possibility of children failing to learn to read by postponing reading instruction until children reach a mental age of 6.5 years, as measured on the Detroit First-Grade Intelligence Test. Overall, studies of this type found that the reason children were having difficulty learning to read was that they were not ready to learn to read when instruction began (Durkin, 1989).

It was during this time that the term "reading readiness" began to appear in the literature (Venezky, 1975). The idea of reading readiness was reflective of the research of the time as well as the philosophy of the educators and psychologists of the period. G. Stanley Hall was one of the leading psychologists at the turn of the century. As the field of child development progressed, Hall exerted a profound influence on what came to be known as reading readiness. In his theory of recapitulation (Hall, 1904), he hypothesized that as individuals develop they pass through the same stages as the human race in its development. All people pass through these stages in a predetermined order that cannot be changed.

According to Hall, performance is determined by heredity. This view was taken one step further by Hall's student Arnold Gesell. Not only did Gesell believe that the course of human development was predetermined, he also proposed that advancement from one stage to the next was a result of what he called "neural ripening" or "automatic and unfolding behavior." He believed that movement from one stage to the next would occur only when an individual had reached a specific level of development, and this could not be accelerated (nor could the order in which these stages were encountered be altered). Movement from one stage to the next occurred naturally as a result of the passage of time (Gesell, 1925, 1928, 1940). This kind of thinking was clearly consistent with views such as those of Morphett and Washburne. Many reading educators felt that children would learn to read only when they had reached the appropriate stage in their development, and that reading instruction should be postponed until then. It is interesting to note that Hall and Gesell did not specify the stage at which reading readiness occurred; however, armed with the convincing data from Morphett and Washburne's work and buoyed by the popularity of the scientific measurement movement, educators pinpointed that stage at a mental age of 6.5.

The idea that a mental age of 6.5 was a necessary prerequisite for learning to read gained widespread acceptance during the 1930s and continued to dominate beginning reading instruction programs for the next 30 years. Objections to the mental age concept of reading readiness began as early as 1936 when Arthur Gates pointed out that the type of reading program is an important determinant of its success, and that the child's mental age is a relatively insignificant factor, correlating about .25 with reading achievement scores at the end of first grade (Gates, 1937; Gates & Bond, 1936; Gates, Bond, & Russell, 1939). Betts (1946) likewise viewed instruction as a viable alternative to waiting. However, objections of this type went virtually unnoticed by educators, who continued to support the use of mental age scores as the indicator of reading readiness.

In a very important sense, readiness tests were developed as an alternative to the mental age philosophy of just waiting. The early reading readiness tests were seen to have several desirable features. They were "designed to measure the traits and achievements of school beginners that contribute to their readiness for first grade instruction" (Metropolitan Readiness Tests, 1933). Not only were they designed to measure overall readiness for beginning reading, but they were also to be used "to predict the rate of development of reading ability, and to diagnose the pupil's status and thus reveal his needs in each of several of the most important abilities required in learning to read" (Gates Reading Readiness Tests, 1939). Presumably, Gates and others (e.g., Betts, 1946) thought that teachers could do something instructionally to promote readiness; they did not have to adhere to a just-wait philosophy. The one feature that readiness tests shared with the mental age approach is that they, like intelligence tests, represented a nonsubjective way to measure a child's readiness for reading, an important criterion in the scientific measurement movement so dominant during this era.

The early readiness tests were paper and pencil, group-administered instruments that focused on vocabulary knowledge, visual discrimination of pictures, objects, letters, or words; auditory discrimination (usually in the form of identifying rhyming words), and copying, which was thought to be indicative of mental maturity as well as physical development (Metropolitan Readiness Tests, 1933). As we have suggested, the intent of these tests was to diagnose individual children's strengths and weaknesses so that they could receive instruction in areas in which they were weak. However, in practice this was not the case. Schools tended to use total scores as indicators of overall readiness for reading instruction; as a result, many children were simply labeled as ready or not ready without any indication of what could be done to help those who were not ready for reading instruction. That was left up to whatever readiness program the school or teacher happened to be using (Durkin, 1989). Tests of this type continued to be developed through the forties and fifties with relatively few changes in format, underlying philosophy, or use.

The 1960s brought changes in both the philosophy of early reading instruction as well as some changes in the readiness tests. The American public was greatly shaken by the Soviet Union's launch of Sputnik I in 1957, and it brought to the foreground public concern over the quality of public school education

in the United States. Not only did people feel that the schools should be teaching more, but that teaching of basic skills should begin earlier (Durkin, 1989). Coincidentally, the thinking of psychologists and educators of the time supported the public's notion that children could and should be taught to read at an earlier age. Among the most prominent and influential thinkers of the time were Benjamin Bloom, Jerome Bruner, and David Ausubel. Bloom (1964) made two important discoveries: (a) preschool children have the ability to learn many kinds of skills, and (b) intelligence develops most rapidly during the first five years of life. Bruner put forth the popular hypothesis that "any subject can be taught effectively in some intellectually honest form to any child at any stage of development" (Bruner, 1960, p. 33). This was supported by Ausubel's definition of readiness as "the adequacy of existing capacity in relation to the demands of a given learning task" (Ausubel, 1959, p. 246). This type of thinking implied that the reason children were having difficulty learning to read lay in the quality of the instructional programs, not in the presence or absence of any predetermined level of readiness or mental age. This view echoes Gates' (1937) comments about readiness in the 1930s. William S. Gray (1969) pointed out that the generally accepted correlation of .65 between mental age and progress in learning to read indicates that there are factors other than mental age that influence learning to read. All of this research stressed the importance of the child's early environment and the necessity of working with the child in order to get him ready to read rather than simply waiting for him to reach an appropriate level of readiness as a result of the passage of time. This kind of thinking, in concert with the social revolution that occurred in the 60s, led to the inception of Head Start programs. These extremely popular preschool programs were designed as intervention programs to help "culturally disadvantaged" children prepare for school and thereby ensure equal opportunities for success in school for all children (Durkin, 1989; Teale & Sulzby, 1986). This effort to teach more to children and to do it earlier was manifested in the schools in two ways. First, the idea that it is necessary to teach prerequisite skills for reading became an integral part of the philosophy of basal reading programs; it took the form of a subskills approach to reading instruction, even in the kindergarten program.. A second outgrowth of this period was that in order to accommodate the teaching of reading in kindergarten, the curriculum was simply pushed back--the skills and materials that had been used at the start of first grade were now part of kindergarten programs. The use of reading readiness programs, usually from basal reading series, became a staple of kindergarten curricula (Teale & Sulzby, 1986).

However, readiness tests of the 60s tended to reflect only a few of the changes in the ways people thought about readiness. The Metropolitan Readiness Tests, for example, continued to be made up of subtests of vocabulary, visual and auditory discrimination, and copying (Hildreth, Griffiths, & McGauvran, 1964, 1966, 1969). There were some exceptions, such as the Murphy-Durrell Diagnostic Reading Readiness Analysis (1964) in which the subtests consisted of phoneme identification (identifying sounds in spoken words), recognition of letter names, and a learning rate test in which the child was tested on the number of words that she could remember after an hour of formal instruction. In general, despite an awakening of programs for early literacy, the readiness tests of the 60s continued the tradition of the previous 30 years.

In the reading field, and in other curricular areas, the 1970s ushered in the era of mastery learning and skills management approaches to curriculum. Ultimately, these movements were to provide a more favorable climate for readiness tests because they, like the readiness tests that had evolved, were based upon a componential assumption about the nature of the reading process. The readiness tests of the 70s included most of the subtests that were found in previous readiness tests--vocabulary knowledge, visual and auditory discrimination--but they were organized in a slightly different way and included some new subtests. Visual and auditory discrimination tasks were subdivided into components. Visual discrimination of printed symbols was tested separately from letter recognition; and rhyming was tested separately from auditory memory (ability to recall a series of spoken words). Subtests of knowledge of letter-sound correspondences appeared consistently on the tests of the 70s in contrast to the 60s, where they appeared only sporadically. The new additions to these readiness tests were subtests on school language (the language of instruction as well as knowledge of standard English) and quantitative language and operations (math). Subtests that required copying were either dropped or rendered optional, mainly because writing was considered to be a skill that followed, not preceded, learning to

read. The standardized readiness tests continued to be designed to permit diagnosis of strengths and weaknesses. In practice, however only total scores were used, and then only as a basis for placing children in groups for basal instruction. Once children began in the basal program, they usually completed it in its entirety regardless of their relative strengths or weaknesses. The driving force for basal instruction tended to be end-of-level tests included in the basal kindergarten programs, which, while uncannily similar to standardized readiness tests, were curriculum specific and, therefore, potentially greater in diagnostic value.

While the readiness tests and classroom instruction of the 70s were most closely aligned with the philosophy of a subskills approach to reading, the research was not. A new approach to looking at beginning reading began in the late 60s and continued into the 80s. Both the maturational and subskills approaches to beginning reading were rejected. In 1966, Marie Clay coined the term *emergent literacy*, which would come to characterize this new approach to beginning reading. The emergent literacy perspective takes into account all of the child's interactions with books, print, reading, and writing as important steps in becoming an independent reader/writer (Teale & Sulzby, 1986). This view also suggests that in a print-rich environment, such as the one in which American children grow up, learning to read and write can be as natural as learning to speak is in language-rich environments (K. Goodman, 1967, 1968; Y. Goodman, 1967; Y. Goodman, 1984; Goodman & Goodman, 1979). The term emergent literacy is appropriate for this movement because children are viewed as being "in the process of becoming literate" through active engagement with their world in reading and writing situations (Teale & Sulzby, 1986). Conversely, the term readiness is inappropriate because it implies that there is a stage in which the child is not yet a reader. Clearly, such an implication is unacceptable to the emergent literacy perspective.

Given the lack of consistency between the dominant philosophies of beginning reading and the formal measures of early literacy we have documented thus far, we should probably not expect to find a match between emergent literacy views and current measures of early reading. To determine whether we will, we now turn to the major task of this study--to analyze the array of tests currently available for evaluating early literacy.

Method

To analyze formal measures of early reading (or readiness), we first had to determine what was available. We started with this operational definition of readiness tests: those tests listed in the reading readiness section of *The Ninth Mental Measurements Yearbook* (Mitchell, 1985). We augmented that definition by calling all of the publishers that we knew published (or suspected might publish) readiness tests. From that combined list, we excluded any tests that were published abroad, for fear that we would not get them in time to meet our deadline. We ordered copies of all the readiness tests uncovered in this search. In this way we received most of the readiness tests readily available to school personnel in the United States today.

Thus, our definition of *formal measure* included commercially available tests and inventories. Even at this early level, we found that most of the *tests* available employ some sort of multiple-choice, machine-scorable format, although a few require students to construct responses and test administrators to make judgments about the correctness, appropriateness, or sophistication of those responses. *Inventories* almost universally take the form of checklists that teachers or evaluators are supposed to use as they observe children's psychomotor, cognitive, social, and emotional development in either natural or controlled settings.

Once we had the tests and inventories, we had to find a way to look at them that would allow us to compare features across tests while depicting the essence of each test fairly. We began with what was essentially a discovery approach, asking ourselves what categories of information we needed in order to characterize each test accurately and to distinguish it from other tests. We listed characteristics we

found, and then we sorted the characteristics into areas that seemed to capture the essence of the tests. These characteristics fell into four broad categories--general test characteristics, skills tested, presentation factors, and response characteristics. Then, each of the four categories was subdivided in order to capture the aspects of each test that were unique to it, as well as those aspects that were common across tests. The subcategories of these four broad categories are listed in Table 1.

[Insert Table 1 about here.]

General Test Characteristics

The category of general test characteristics focused on the aspects of the tests that were basically descriptive of tests in general--copyright, age, administration, time, format, number of items, and stimulus size. While all of the tests used in this analysis are currently in print, the copyright dates ranged from 1976-1988. Whenever more than two copyrights were still being marketed, only those versions with the most current copyright date were included in our analysis.

Age refers to the age of the child for which the publisher recommends administration of the test to be appropriate. *Administration* refers to the setting in which the test is designed to be used, individual or group. All of the group tests could be given to children individually, but the individual tests are not appropriate for use with more than one child at a time. *Time* indicates the amount of time that the publisher or author of the test recommends for administration of that test. In terms of *format*, a test could either have a multiple-choice format or a checklist format. In a checklist format, the examiner was usually directed to check off the appropriateness or occurrence of student responses to test items on a predetermined list.

The last two subcategories were the number of items used to test each skill and the size of the stimuli. The *number of items* used to test a child's knowledge of individual skills ranged from 1 to 54. The *size of the stimuli* that the child responds to was also categorized. The stimuli ranged from 1/8 inch to 10 inches, with almost 80% of the stimuli measuring 1 inch high or less.

Skills Tested

In this category each subtest was classified according to the skill that was being assessed. These skills fell into five major areas--world knowledge, symbol and sound concepts, literacy and language concepts, comprehension, and others.

1. *World knowledge* includes the specific skills of vocabulary, picture identification, and sequencing. Regardless of the specific skill label, it is most often assessed by asking the child to identify pictures that represent general vocabulary or conceptual knowledge about everyday things or occurrences. For example, an item of this type might ask the child to "mark the picture of the truck that came to put out the fire" (see Figure 1, example a) or "mark the picture of the fan" (see Figure 1, example b).

[Insert Figure 1 about here.]

2. *Symbol and sound concepts* includes most of what we would normally label visual discrimination, auditory discrimination, letter name and letter sound knowledge, and a few outliers such as spelling, syllable knowledge, and memory. A child's knowledge of letters and sounds is assessed in a variety of ways including visual matching of letters, auditory discrimination of words with similar parts, matching letters with pictures containing the sound of the letter, recognition of letters or sounds, identification of letters or sounds, and production of letters or sounds. For the purposes of illustration the letter *b* will be used to exemplify these different types of item formats (see Figure 2).

[Insert Figure 2 about here.]

3. *Literacy and language concepts* includes items that assess a child's knowledge of how to use language, reading strategies and conventions of printed language, word recognition, and the language of instruction. Language usage was assessed using items that tapped the child's knowledge of syntax and grammar (see Figure 3). The assessment of children's understanding of reading strategies and the conventions of printed language included items that asked children to demonstrate their ability to do tasks like identifying where to begin reading a book, demonstrating left-to-right and top-to-bottom progression, identifying the difference between letters, words, punctuation, and pictures, as well as knowing that the print carries the message. The third area in this major skill category was the child's knowledge of school language. Items of this type looked at the child's knowledge of relationships such as over, under, inside, around, first, next, last, and so forth. Mainly because we could not find a more appropriate major skill slot in which to place them, we included a fourth type of subtest in this category--word recognition. Most often, this skill was tested by providing children with a set of four words, one of which is to be circled when the teacher gives a directive like, "Circle the word that says GO." Occasionally, the children were asked to "read" the words. In all cases, the words on these tests were high-frequency words, the kind that dominate most basal preprimers.

[Insert Figure 3 about here.]

4. *Comprehension* assessment took various forms. The most common was one in which the examiner read a short story of a few sentences to the child, who was then to choose the picture that best represented what had been read (see Figure 4, example a). A similar form was one in which the examiner again read a short story to the child and then asked a question about what had been read. The child was then to choose the picture that best answered the question (see Figure 4, example b). A less-common form was one in which the child was asked to actually read a sentence or short story and to fill in a missing word or answer a question about what had been read (see Figure 4, example c).

[Insert Figure 4 about here.]

5. *Others* consisted of miscellaneous areas that did not fit into any of the other areas. Included in this category were items that assessed the child's knowledge of shapes, colors, mathematics, and quantitative language (see Figure 5). These types of items were excluded from the other major skill categories because they are not directly related to reading readiness even though they may be related to school readiness in general.

[Insert Figure 5 about here.]

Presentation

In this category we looked at the ways the items were presented to the students. Two subcategories of the presentation of items were considered--mode and unit. In the *mode* of presentation, each item was classified in terms of the primary mode of presentation used by the examiner. The modes of presentation used on readiness tests were auditory, visual, and written. An item that was presented using an *auditory* mode was one in which the child was to respond to something that the examiner *said*. For example, "Mark the picture of the leaf". Items that were presented using a *visual* mode required that the child respond to something that was *printed* in the test book--usually a letter, word, or picture. The final mode, *written*, required the child to respond to something that the examiner *wrote* on the spot, such as asking the child to copy his name after the examiner had written it.

The *unit* of presentation refers to the type of stimulus that the child was asked to respond to--graphemic, phonemic, syllabic, word, phrase, sentence, connected discourse, patterns, numerical, picture, or book. Most of these categories are transparent in that the label defines each; however, there is potential overlap among some of them. For example, often the teacher *said* "words" in both the phonemic and word categories. However in the phonemic category, the focus would have been on finding a picture

that starts with the same sound as, for example, *candle*, while in the word category, the focus would have been on the lexical item itself as, for example, in finding the picture of the candle. See Figure 6 for examples of these various levels of unit of presentation.

[Insert Figure 6 about here.]

Response

This category of children's responses has three subcategories--level of processing, stimulus type, and student response mode. The *level of processing* required of the student to respond correctly to an item ranged from recognition to identification to production. A *recognition* item was one that required the child to select the correct response from a list of alternatives such as, "find the letter *a*" when the child is looking at the group of letters "d o a p." In an *identification* item, the child would be required to name the letter; and in a *production* item, the child was asked to write the letter.

Unit of response looked at the stimuli that the child was to use to indicate the correct answer to the item. The units of response were picture, letter, word, nonsense word, phrase, sentence, connected discourse, book, numerical, or objects. See Figure 6 for examples of the levels of stimulus type.

The *student response mode* was simply a categorization of what the student had to do in order to respond to the item. The response modes were underlining, marking, pointing, filling in the circle (or bubble), writing, drawing, manipulating, and oral responses.

Results

We report our results in two distinct sections and methodological traditions. First, we provide a set of short "case studies" for several of the tests in our corpus. For example, we review the Metropolitan Readiness Tests (MRT) (Hildreth, Griffiths, & McGauvran, 1964, 1966, 1969) in some detail because of widespread use and of their "representativeness" among group tests. But we also examine a few of the "outliers," such as Marie Clay's Concepts About Print (CAP) (1979) assessment, in order to provide a sense of where the boundaries of our corpus of tests lay. Second, we provide a quantitative perspective on our corpus by examining the descriptive statistics generated from the analyses detailed in our methods section. We opted for descriptive rather than inferential statistics because we found that most effects worth discussing were obvious rather than subtle.

Representative Examples

The Prototype

In order to get a feeling for make-up of the readiness tests that are available we decided to describe the test that seems to exert the greatest impact on the field, the MRT. It was chosen because it is the most widely used group test. The MRT has two levels. Level 1 is designed for use with children who are at the end of prekindergarten through the middle of the kindergarten year, and Level 2 is for children in the middle of kindergarten through the beginning of first grade.

Level 1 begins with a test called *Auditory Memory*. In the first section of this test, the teacher first identifies all the pictures that will be used on the test (spoon, tree, chair, cat, house, ball, star, cup). Then the teacher says the names of three objects, such as *chair, house, tree*, and the child is to mark the box that contains a picture of those objects in the same order that the teacher read them (see Figure 7). There are 12 items of this type.

[Insert Figure 7 about here.]

In the next section, labeled *Beginning Consonants*, the children are to find a picture of the word that begins with the same sound as the two words the teacher says. It really represents an auditory discrimination task except the focus is always on matching the target sound. Perhaps auditory matching would be the best label. No association with letters is required at all. For example, the child looks at pictures of a cat, a sock, milk, and an arm, and is to choose the one that begins with the same sound as *miss* and *mine* (see Figure 8). There are 14 items like this.

[Insert Figure 8 about here.]

The third section of the test involves *Letter Recognition*. The child looks at four letters and marks the one that the teacher names. Four of the items use upper case letters and six use lower case letters (see Figure 9).

[Insert Figure 9 about here.]

In the fourth section, *Visual Matching*, the child looks at a target pattern, letter, group of letters or numbers and then marks the box that has the same thing as the first (i.e., target) box (see Figure 10). In the actual student booklet, the target box is highlighted in a different color. The test has 14 visual matching items.

[Insert Figure 10 about here.]

The *School Language and Listening* section of the test has 15 items that test the child's knowledge of concepts and language used in school (no, between, center, cut, some, carrying, laughing, giving, flying, next to, from, indoors, larger, bigger, smaller, exactly the same). The child is to mark the picture that represents what the teacher says (see Figure 11).

[Insert Figure 11 about here.]

The last part, labeled *Quantitative Language*, has 11 items that are related to math, such as counting, numerals, ordinal position, etc. Again, the child marks the picture that represents what the teacher says, "Mark the second duck" (see Figure 12).

[Insert Figure 12 about here.]

The Level 2 test is similar to the Level 1 test in many ways. The same basic categories are represented, but the items are considerably more difficult. Also, some of the types of behaviors tested on a single test in Level 1 are broken into two tests in Level 2. For example, there are separate tests for *School Language and Listening* in Level 2, and the quantitative section is broken into *Quantitative Concepts* and *Quantitative Operations*. *Visual Matching* from Level 1 is augmented by a new variation called *Finding Patterns*, in which a student now has to find the target pattern within a visually more complex array (e.g., Find the *it* in *visit*) The entire section on auditory memory is dropped and a completely new section is added--*Sound-Letter Correspondences*, in which the children are to choose the letter, from among four letters, that has the sound heard at the beginning of the word represented by the picture at the beginning of the row. For example, the children look at a picture of a seal and are then to mark the *s* (see Figure 13).

[Insert Figure 13 about here.]

As we suggested, the other sections of the Level 2 test differ from the Level 1 test in the difficulty of the items. For example, on the Level 1 test, the children are asked to match patterns that are up to three letters long, and on the Level 2 test they must match patterns that are up to six letters long; further, on the *Finding Patterns* test, the patterns are embedded in another, more complex, pattern.

Our examination of the other tests within our corpus leads us to the conclusion that the MRT is typical--wholly representative of items that are found on the other group-administered readiness test. The other tests provide more variations on a theme than they do completely different "compositions."

When used in conjunction with the ancillary materials that either accompany or can be purchased with the MRT, it is apparently designed to be used for two purposes: (a) diagnosing students' strengths and weaknesses in the areas tested, and (b) helping teachers decide how to group children for instruction. We use the qualifier apparently because we could find no passage in the examiner's manual in which possible uses of the test were laid out unambiguously. But, by virtue of the fact that the class record sheet provides subtest scores for individuals and suggests placing students into one of three groups (plus, check, or minus) based upon subtest scores, we inferred that the authors intend it to be used for diagnosis and for grouping. The MRT also provides a form to be given to parents that explains the tests, shows how their children performed on the tests, and makes suggestions about things the parents can do to help their children improve.

Outliers

If the MRT represents the center of our distribution of readiness tests, then the three tests we describe in this section represent the mavericks, the outliers that account for the variability within our data set.

Developmental checklists. One type of readiness test that does not fit into a classification system with the tests already described is the developmental checklists, which may either accompany the standard group test or stand on its own. These checklists are designed to be used with individual children. They look at the child's development in areas such as social-emotional development, physical development, cognitive development, and language development (see Table 2 for a more complete list).

[Insert Table 2 about here.]

In order to use one of these checklists, the parent or teacher simply observes the child and checks off the behaviors or characteristics the child exhibits. When the checklist accompanies a standardized group test, as in the case of the MRT, it is recommended that the information from the checklist be used as an additional source of information to aid in the interpretation of the results of the group test. The stand-alone checklists (e.g. Humanics National Child Assessment Form) are designed to chart individual growth and development over time and as an aid in planning educational experiences for the child. In a sense, such checklists become convenient "reminders" for teachers; however, the basic responsibility for making judgments about student progress rests squarely on the shoulders of the teacher.

The Gesell School Readiness Test. Another test that did not fit into our original classification scheme was The Gesell School Readiness Test. We discuss it only because it has such a long history of use as a device for helping school personnel make decisions such as early kindergarten entry and referral to special programs. The primary emphases of this test are conceptual knowledge and perceptual-motor performance. Conceptual knowledge items involve tasks like identifying body parts or talking about the world in which the child lives (e.g., "What do you like to do best at home?"). The perceptual-motor tasks involve copying shapes, writing letters and numbers, and the Cube Test. The examiner also notes physical characteristics and behaviors. The Gesell Test is intended to measure the child's "developmental age," a construct purposely designed to differ from chronological age or experience. In this regard it is recommended for use as a screening device for placement in special programs or for delayed entry into ordinary programs (Gesell Institute, 1987). Other evidence (e.g. Shepard & Smith, 1986) confirms its use in exactly these ways.

The Gesell Test has been criticized on a number of grounds--its outdated theory of child development (Meisels, 1987), lack of reliability and validity evidence (Carlson, 1985; Meisels, 1987), lack of a usable scoring system (Carlson, 1985), an unverified construct for developmental age (Meisels, 1987), and

wholescale misapplication in school settings (Meisels, 1987; Shepard & Smith, 1986). One can only conclude that its use is fraught with controversy, and that anyone considering its adoption should be prepared to enter the fray.

The Concepts About Print (CAP) tests. Another outlier we ended up including in our corpus, even though it was not reviewed in *the Ninth Mental Measurements Yearbook*, was this test developed by Clay. The CAP is really part of a larger battery of observational devices known as the Diagnostic Survey (Clay, 1985), which includes other "reading" assessments, such as letter identification, word identification, writing vocabulary, and running records of the child's oral reading of connected text. In New Zealand, the CAP tests are not used as readiness tests, instead they are used only as a "progress check" one year after instruction has begun for those children identified as potential "problem" readers. In the United States, however, the CAP tests have been used in kindergarten in much the same way the readiness tests are often used; hence we have included them in our corpus.

What distinguishes the CAP from other readiness tests is that it is administrated in the context of "reading" a real book. Each five-year-old sits with the examiner and is asked to help by pointing to certain features as the book is read by the examiner. For example, the examiner would ask, "Where do I start reading?" or "Find a little letter like this" (the child is shown upper case T) or "Show me a word," and the like. The CAP is used in the United States as an aid to the teacher about one aspect of learning during the early stages of reading acquisition--knowledge of significant concepts about printed language. It should be noted in passing that in New Zealand, where the CAP is used as a part of a progress check, *no* systematic assessment of children is done prior to the onset of reading instruction.

Analysis of the Corpus

We report our results for all the types of categories used to analyze the tests we examined. Typically, we used one of two dependent measures--either the *number of subtests* exhibiting a particular feature or the *number of items* in a certain category. To assist the reader, we have organized our data reporting and discussion around three broad questions:

- How are the tests administered? (issues of administration, time, and format),
- What kinds of things are tested? (issues of the type of major and minor skill categories assessed), and
- How are students tested? (issues of item presentation and student response).

How Are the Tests Administered?

Of the 20 tests in our corpus of currently available reading readiness tests, 16 (80%) were group tests and 4 were individual tests. Aggregated at the level of subtest, of the 208 subtests we examined, 170 (82%) were in group tests while 38 were in individual tests (18%). So, individual and group tests tend to assess just about the same number ($n = 10$) of subtests per test.

There is complete covariation between type of administration (group versus individual) and test format (multiple-choice versus checklist), with all group tests using a multiple-choice format and all of the individual tests using a checklist format. By checklist format, we mean that the test authors provide a checklist for each child, with spaces to record a test score and anecdotal comments for each subtest or, in some cases, for each item. Additionally, we found that three of the "tests" we ordered were not really tests. Instead they were developmental checklists that provided a long list of student behaviors and strategies that ought to be observed and judged; however no explicit strategies were offered for conducting the observations. We decided not to include these three developmental checklists in our analysis.

The recommended ages for use of readiness tests ranged from prekindergarten through the beginning of fourth grade, however, 87% of the tests were recommended for use between the beginning of kindergarten through the middle of first grade.

While one-third of the tests did not come with recommended time guidelines, the time needed for administration of the remaining two-thirds of the tests ranged from 2 minutes to 225 minutes, with an average test time of 127.38 minutes (sd=75.52). However this overall mean is misleading because of the huge difference between group and individual tests. Group tests averaged 162.4 minutes (sd=45.67), while individual tests average 10.67 (sd=7.36). These huge discrepancies bring into question the widespread assumption that group tests are necessarily more efficient than individual tests. Notice that a teacher could test 16 children individually in the time necessary to administer one group test. And this figure does not take into account either practice time or scoring time! One might argue that the teacher who gives a group test is getting lots more information on *more* subskills or aspects of reading readiness, thus weakening our argument against the inefficiency of group measures. But such an argument is invalid on several counts. First, the information from the group tests is rarely very curriculum specific. So a teacher learns that Georgette knows half the alphabet, but not which specific letters; and the 50% inference will likely be based upon a sample of 4-10 letters. Second, many of the skills tested on group tests, while they may have reasonable predictive validity (they'll sort out the good readers from the poor), will not necessarily ever become part of a curriculum. Auditory memory tests are a good case in point; short of helping students learn more about their world, there is nothing to teach. Third, for very young children, knowledge of directions and task requirements have to play a major role in group test performance; the tasks are simply too novel for too many young children to permit any other inferences.

What Kinds of Things Are Tested?

There are several ways to answer the question about what is tested--by individual skill, by major skill category, for each test, for each type of test. We will try to balance all of these perspectives in reporting our results. In Table 3, we report the average number of items per subtest and the number of subtests by major skill category for each of the 20 tests included in our corpus. The first conclusion to draw from these data is that variability characterizes the data set. Most tests test a little bit of everything, but some, such as the Boehm R or the tests developed by Clay (Concepts about Print and Ready to Read), are very specialized. Variability notwithstanding, the summary data are of interest. The last line in Table 3 (percent of total items) is a simple calculation of the percentage of items from the entire corpus of items (n=2,652) that assess that type of knowledge. One-third of the items (but 48% of the subtests and, therefore, 48% of the judgments) are devoted to symbol-sound knowledge, with the second largest chunk devoted to literacy and language concepts (22% of the items and 21% of the subtests). The mean number of items by major skill category data merits the interpretation that, on average, test constructors devote fewer items per subtest (presumably the unit from which diagnostic judgments would be made) to symbol-sound tests than they do to other types of tests.

[Insert Table 3 about here.]

There is, of course, much variability within each of these five major skill categories. Table 4 depicts that variability by reporting the number and percentage of subtests devoted to each of the specific skills from the entire corpus. For example, the 48% of the corpus figure for symbol-sound knowledge covers a multitude of sins, ranging from simple visual discrimination tasks to more difficult auditory tasks to letter knowledge to knowledge of specific letter-sound correspondences; most of these "sins" are well-represented in the corpus.

[Insert Table 4 about here.]

How Are Students Tested?

The answer to this question can be either simple or complex. To get a broad picture of how students are tested, we first examine each of the relevant variables individually. We can say with some confidence that, in terms of level of processing, students are rarely asked to produce or construct answers to test items. We know that of the 208 subtests in our sample, 148 (72%) require students to recognize a response (e.g., find the letter A in this group), 49 (23%) require identification (e.g., what is this letter I am pointing to), and only 11 (5%) invite production (e.g., write the letter A). Teachers, or test administrators, are very clearly in control of the situation.

In Table 5 we look at the issue of level of processing from a slightly different slant. The data in Table 5 are the average number of items per subtest and the number of subtests broken down by major skill category, type of administration, and level of processing. These data suggest that the heavy reliance on recognition activities is primarily due to the preponderance of group tests in the readiness marketplace. Notice that identification and production are much more common in individual tests.

[Insert Table 5 about here.]

The data for unit of presentation and unit of response are reported jointly in Table 6. The row and column frequencies and percentages provide data for each measure independently; the cell totals provide an interesting picture of the natural covariation between units of presentation and units of response. The summary data for the last two columns (frequency and percentage for units of presentation) corroborate what every novice, let alone experienced, kindergarten teacher knows: Students respond to a lot of pictures--marking them, circling them, or filling in the bubble next to them. Test makers (and workbook makers) rely heavily on pictures in creating their materials. Looking down those two columns, it is clear that students also respond to letters and words--marking them, circling them, or filling in the bubble next to them. In fact, those three stimuli--pictures, letters, and words--account for 81% of all of the units of response in our corpus. Units of presentation tend to be somewhat more varied than units of response, with graphemic units, phonemic units, words, and sentences represented about equally in the corpus.

[Insert Table 6 about here.]

One note of explanation about the *phonemic unit* of presentation category is necessary. In over half of the *phonemic* cases, the test administrator actually said a "word"; we decided to classify these cases as phonemic rather than word because the clear intent of the item was to focus on a sound, as the auditory discrimination task, "Find the picture that starts like ball," or the letter sound task, "Find the word that starts like ball." Those "word" units of presentation that were classified as word tended to be vocabulary items, as in, "Find the picture of the dog."

It is interesting to note the natural covariation between units of response and units of presentation. For example, almost all *graphemic* units of presentation are associated with *letter* units of response. *Sentence* units of presentation are followed by *picture* responses about two-thirds of the time; typically, these are picture comprehension items in which the administrator says something like, "Find the picture of mother ironing Mary's dress." *Phonemic* stimuli were most often followed by pictures (mostly auditory discrimination) but sometimes by either letters or words (in both cases, letter-sound knowledge is normally being tested).

Table 7 contains the mundane, but interesting data on how children are supposed to indicate their answers to test items. What is most notable about these data is the dominance of the fill-in-the-circle (or bubble or oval) activity. As with the data on level of cognitive processing, it appears that these data result from the heavy reliance on machine-scorable, group administered tests. Equally remarkable is

how seldom children are asked to do anything resembling an active mode of responding, such as saying something (11.5%) or writing something (3.4%).

[Insert Table 7 about here.]

Finally, Table 8 presents data on the mode of presentation of the stimuli for test items. Almost 75% of the time, children are directed to listen to something a teacher or test administrator says. Occasionally, they respond directly to something they see (these are almost all letter or pattern matching activities). Only twice in our entire sample were they to respond to something a teacher was supposed to write on the spot.

[Insert Table 8 about here.]

Summary of Results

In summary then, we can say that readiness tests, on the whole, have these prototypic characteristics:

1. They are typically administered to a group of children, rather than to an individual child.
2. The clear emphasis (almost half of the subtests) is on sound-symbol knowledge.
3. They take a long time to give (a little over two hours spread out over a few days).
4. They rely a lot on students filling in bubbles based upon what the teacher says or what the students see in a picture.
5. Recognition, not identification and production (the staples of reading), is the dominant level of cognitive processing involved; the group testing situation seems to be responsible for the dominance of recognition activities.

Discussion

Portrait of a Kindergartner

One way of getting a handle on the operational definition of early reading emanating from readiness tests is to look at what happens during one such test from the point of view of a five-year-old child. What does the world of reading look like to a typical kindergartner, Sam, taking a reading readiness test? First, he is told to go to bed early the night before the test and eat a good breakfast before coming to school on the day of the test. On the first day of testing, Sam sits at his place at his table with books set up all around him to make sure that no one else can look at his paper. He is given a sharp new pencil along with the instructions that it not to be used to write on the test except for filling in the bubbles when the teacher tells him to. Next, the children practice filling in bubbles. The bubble must be completely filled in using marks that are very dark and they must not go into any other bubble. After learning how to fill in bubbles properly, Sam is told that he is to erase the marks he just made, you see he needs to practice erasing completely just in case he wants to change his mind about something he does on the test. The next thing to learn about is the marker--how to put it under the row with the picture of the star at the beginning. This, Sam learns, is how he is to know which pictures he should be looking at.

Now the class is ready to begin to take the test. The teacher tells them that everyone should work very hard and try to do their best because the test will tell their parents and the first-grade teachers how much they have learned in kindergarten. They are also told not to talk or look at anyone else's test.

Sam puts his marker under the picture of the star and is ready to begin, but Sara can't find the star and Jim dropped his pencil on the floor. Finally, everyone is ready to begin the reading test. They find pictures, listen to sentences, find letters, and count things in pictures. When they are all finished, the teacher explains that what they have just done is the practice reading test; the real one will begin tomorrow and will look just like the one they did today. The class will be working on the test all week. Over the next few days, Sam and his classmates fill in lots of bubbles. Bubbles that tell which group of letters and which shapes look like the thing in the box at the beginning of the line. Bubbles that tell which picture begins with the first sound in the word they hear, bubbles that tell which letter the teacher says, and bubbles that tell which picture goes with the word or sentence the teacher says. Sam wonders when they are going get to read.

While this scenario may seem a little harsh and overstated, it is more typical than we would like to think. In the name of objectivity and efficiency, children are receiving odd messages about the nature of literacy. It is important to look at the readiness tests in terms of their usefulness to school personnel, but it is equally, if not more, important to look at the inferences about the nature of reading that children and their parents will draw from these sorts of tests.

What Have We Learned?

In looking at the data from the analysis of the readiness tests, we have uncovered several facts and generalizations about them, all equally unsurprising to us and to educators.

Overall, readiness tests in our corpus look very much like those designed by Gates and others in the 1930s; furthermore, they look very much like one another. Most of these tests are group tests presumably because of the assumption that it is more efficient to test a group of children all at one time than it is to test them individually. As we pointed out, this practice may rest on a faulty assumption about efficiency; it takes an average of 162.4 minutes to administer a group test (not including practice or scoring time), while it only takes an average of 10.67 minutes to administer an individual test. Therefore, once you have taken into account the additional time it takes to administer the practice group tests and score them (not to mention practice time), it may be just as efficient to administer individual tests; and, as we have argued, the information obtained from individual tests is often more useful to the teacher than are scores from norm-referenced tests.

An equally unstartling finding was the heavy emphasis on sound-symbol knowledge. Almost half of the subtests dealt with sounds and symbols. This overwhelming emphasis on sound-symbol knowledge sends a very clear message--knowledge of sounds and symbols is a primary goal of early reading.

In terms of the ways children are tested on formal measures of early literacy, most are, to say the least, disconcerting. Children are tested on isolated skills in decontextualized settings, rather than on reading tasks in situations in which they are asked to behave like readers. This conclusion is supported by several pieces of data. First, the examples presented to illustrate skills tested portray a picture of isolated, not integrated, reading skills. Second, recognition, not production or even identification, dominates as the primary mode of cognitive processing. One wonders what happened to the theory of reading as a constructive process. Third, when they recognize things, children are usually asked to respond either to a picture or to something the teacher says. At the very least, real reading involves identification of words in sentences. Finally, what dominates the whole enterprise when children actually take the test is test-taking behavior--filling in bubbles, moving the marker, making sure everyone is in the right place. These activities may be related to test taking, but they have nothing to do with reading.

An important final point to consider in regard to these formal measures of early literacy is the message they are sending about what really counts. If you take the tests to be a reflection of the field's priorities in early literacy instruction, then what matters most is children's ability to recognize pictures, letters, and sounds so they will be ready, someday, for the real thing.

Beyond Standardized Readiness Tests

We chose to focus our discussion of formal measures of early literacy on readiness tests that were available as stand-alone entities on the professional market. We did not include some other candidate indices of early literacy, such as the tests accompanying readiness levels of basal reading programs or the levels of standardized reading tests intended for first grade. Consequently, the question of whether we have fairly represented the range of early literacy measures can be raised. In order to answer this question we now turn to a brief review of some of these measures of early literacy to see if they paint a different picture from the one we have painted thus far. We look at tests from basal readiness programs, some early levels (first grade) of standardized tests, and first-grade basal tests.

Basal readiness tests. On the grounds that the tests that they provide for measuring mastery of the program provide powerful evidence of what they consider important at this level, we looked at several of the leading basal series--Holt, Rinehart & Winston; Houghton Mifflin; Laidlaw Brothers; Macmillan Publishing Co.; and Silver Burdett & Ginn--to see how they assessed student mastery of their programs. In looking at these tests we were struck by two things. First, these tests are remarkably similar to one another. In all cases the children are tested on letter recognition, auditory discrimination, and visual discrimination (in the form of matching letters and/or words). While there are some differences between the tests--sight words appear on two of the tests, and knowledge of vowel sounds and rhyming each appears once--these tests are more alike than they are different from one another. Second, the degree of similarity between the basal tests and standardized tests, such as the Metropolitan Readiness Tests, is as remarkable as their similarity to one another. Not only are the skills tested similar, but the format is nearly identical. For example, notice the similarity between the way letter sound knowledge is assessed in Figure 14 (extrapolated from one of our basals) and the format in Figure 13 (extrapolated from the MRT). In making these comparisons, we were tempted to conclude that the basal publishers had the standardized readiness tests in mind (if not in hand) when they constructed their own readiness tests. This is not surprising in view of the fact that many schools use standardized tests as an indicator of the effectiveness of their instructional programs; and one way to help the students do well on the standardized tests is to give them practice on similar activities, which the basal tests provide. In sum, then, the first alternative picture of early literacy we have examined suggests that our first picture was representative.

[Insert Figure 14 about here.]

First-grade standardized tests. Next, we turn to standardized reading tests designed to be used with first-grade children. These tests generally include subtests on vocabulary, word recognition, language usage, and reading comprehension. The vocabulary section is made up of items in which the student is to choose the word that best completes a sentence. For example, Mike _____ to school today. (*went*, *likes*, *home*, *fast*). In word recognition items, the student is to select the word that begins or ends or has the same vowel sound as the word the teacher says--"Find the word that begins with the same sound as boy." (the child looks at *call*, *work*, *bat*, *my*). Language usage items involve having the child choose the word that best completes a sentence. For example, "The _____ are playing." (*girl*, *girls*). The reading comprehension subtests are made up of two types of items. In the first, the child reads a sentence and chooses the picture that shows what the sentence is saying. In the second type of reading comprehension item, the child reads a short story of a few sentences and then answers a question or two about it.

First-grade basal tests. Like the readiness tests, the first-grade basal tests are remarkably similar to comparable levels of the standardized tests. Typically, they include tests of word recognition, letter-sound correspondences, and reading comprehension. In word recognition items, the student is to mark the word the teacher says. For example, the student looks at three words--*with*, *you*, *in*--and is to told to mark the word *with*. For letter-sound correspondence items, the student is to mark the letter or word that has the same beginning, ending, or vowel sound as the word the teacher says.

Among basal tests can be found the same two types of comprehension item that we described for the standardized tests.

While the first-grade tests, both basal and standardized tests, paint a somewhat different picture of early literacy than the readiness tests, there is still a considerable amount of overlap between them. They all subscribe to the same philosophy--a subskills approach to reading instruction. In one sense, the first-grade tests simply represent the next logical step following the readiness tests.

In answer to the question of fairness in representation we raised earlier, we feel confident that the similarity we have noted between formal readiness measures and those measures found in basals and on standardized reading tests permits the conclusion that we have provided a fair representation of early literacy assessment in our discussion of the data.

Working Toward Alternatives

Why do we have these types of tests? How did we get into a situation in which important judgments about readiness for reading are based upon data from tests that, at best, only predict reading success? What are the alternatives we could consider? These are the questions with which we conclude our discussion of early literacy measures.

It is our conclusion that the measures of early literacy that dominate our field result from the same philosophy and traditions that undergird reading assessment generally. The assessment (and, no doubt, the instruction) of both early literacy and advanced literacy are beset by implicit, if not explicit adherence to three related, counterproductive, and fallacious constructs--decomposition, decontextualization, and objectivity.

The subskills approach to reading instruction and assessment is based on the fallacy that the way to make a complex task, like reading, simple is to decompose it into smaller parts; then each part can be worked on until it is mastered; finally, the parts can be resynthesized into a working whole.

Once reading has been broken down into its "component parts," decontextualization takes over. Each of the "component parts" is then practiced in isolation until students achieve mastery. The assumption is that by isolating the skill in question, the student can focus directly on what needs to be done to master it. If the skill were practiced in a real reading context, so the argument goes, the extra contextual support would only divert the students' attention from the real task at hand--to learn the skill--and, hence, cloud the issue.

It is on these assumptions that most formal measures of early literacy are based. They test the "component parts" of reading in isolation from any natural, real reading context to see whether they have been mastered. These formal measures of early literacy provide schools with an added attraction nurtured, if not driven, by the accountability movement in educational politics. They give the illusion of being *objective*. Important decisions about children can be made on the basis of tests rather than, so the argument goes, on the basis of subjective judgments of fallible human beings. The irony here is that the tests are no more objective than teacher judgment; they simply move the subjectivity one level farther away from the student--to the fallible human beings who construct the tests in the first place.

What are the alternatives? What is the solution to the dilemma imposed by having formal measures of tasks few would label literacy? While we recognize all the problems attendant to individual assessment--most notably, administrator bias, inefficiency, lack of normative information, and subjectivity, we want to come down clearly on the side of what some (Collins, Brown, & Newman, 1989; Valencia, Pearson, & McGinley, in press) have come to call "situated" assessments of individual students. We are convinced, based at least in part on our dissatisfaction with what is currently available, that the most productive methods of assessing early literacy lie in individual assessment of children while they are

engaged in a literacy activity. Situated assessments are desirable for two reasons. The first is that they maximize the likelihood of optimal performance rather than typical performance since the child has more "data" and cueing systems available for rendering a task sensible. Second, since situated assessments arise from the instructional setting, the issue of generalizability to real instructional situations disappears because there is no gap between assessment and instruction. Of the tests we have reviewed, only those in Clay's battery come close to meeting such a standard. Clearly, we need more work in this area so that we can develop approaches that are even more situated than Clay's. By the way, we would not go so far as to say that knowledge of specific skills, such as symbol-sound correspondences, are unimportant; but we would insist that they be dealt with in context, as needed and used by children engaged in the task of trying to render print sensible.

The major purpose of formal measures of early literacy is to help teachers make informed decisions about issues such as what materials to select, how to group students for instruction, and what students need to practice. We acknowledge the fact that tentative initial decisions like these can be made (indeed are made) on the basis of standardized readiness tests or basal tests. But, we all know that learning and teaching are processes that are too dynamic to rely upon single measures of anything, especially early literacy. More than better tests, teachers need more knowledge. They need the kind and depth of knowledge about reading that will permit them to develop the expertise to construct situated measures that match their needs and their students' needs very closely. Some of these situated measures will have to be constructed "on the fly" while trying, for example, to answer questions like: Will Henry Smith stand half a chance of reading the library book he has chosen? Others will be built for later reflection in trying to answer questions like: Are the students in my class becoming more reflective readers? But all will relate directly to students, tasks, and texts in a particular situation.

To prove that we can write with a forked tongue, we admit to a conviction that a second part of the solution lies in the construction of better commercially available tests. We adhere to this conviction because we have difficulty envisioning a future in which politicians and administrators learned to trust our judgments completely; for better or worse, there will probably always be a place for wide-scale assessment of most things that matter to us.

But whatever new tests or procedures we build, we must insist that they reflect not only the emerging tradition in emergent literacy but also the fact that teachers are professionals who bear the ultimate responsibility for making informed decisions. It is our hope that those decisions will be based on the very best information we can gather. Also, no matter how objective and detached we try to become in our assessment endeavors, these endeavors will always involve professional judgments, even if those judgments are moved out of the scoring and interpretation stages to the stages of conceptualization and item development. Hence the quality of the information we get from assessment activities will vary directly as a function of the knowledge that educators bring to those activities. It is increased awareness and knowledge through education that should be the focus of our efforts to reform literacy assessment.

References

Ausubel, D. P. (1959). Viewpoints from related disciplines: Human growth and development. *Teachers College Record, 60*, 245-254.

Betts, E. A. (1946). *Foundations of reading instruction*. New York: America Book.

Bloom, B. S. (1964). *Stability and change in human characteristics*. New York: Wiley.

Bruner, J. (1960). *The process of education*. Cambridge: Harvard University Press.

Carlson, R. D. (1985). Gesell school readiness test. In D. Keyser & R.C. Sweetland (Eds.), *Test Critiques Vol II* (pp. 314-318). Kansas City, MO: Test Corporation of America.

Clay, M. M. (1966). *Emergent reading behavior*. Unpublished doctoral dissertation, University of Auckland, New Zealand.

Clay, M. M. (1985). *The early detection of reading difficulties* (3rd ed.). Auckland, New Zealand: Heinemann.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Erlbaum.

Dickson, V. E. (1920). What first grade children can do in school as related to what is shown by mental tests. *Journal of Educational Research, 2*, 475-480.

Durkin, D. (1989). *Teaching Them to Read* (5th ed.). Boston: Allyn & Bacon.

Gates, A. I. (1937). The necessary mental age for beginning reading. *Elementary School Journal, 37*, 497-508.

Gates, A. I., & Bond, G. L. (1936). Reading readiness: A study of factors determining success and failure in beginning reading. *Teachers College Record, 37*, 679-685.

Gates, A. I., Bond, G. L., & Russell, D. H. (1939). *Methods of determining reading readiness*. New York: Bureau of Publications, Teachers College, Columbia University.

Gates Reading Readiness Tests (1939). Teachers College, Columbia University.

Gesell, A. (1925). *The mental growth of the preschool child*. New York: Macmillan.

Gesell, A. (1928). *Infancy and human growth*. New York: Macmillan.

Gesell, A. (1940). *The first five years of life*. New York: Harper & Bros.

Gesell Institute (1987). The Gesell Institute responds. *Young Children, 42*(2).

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 4*, 126-135.

Goodman, K. S. (1968). *Study of children's behavior while reading orally*. (Final Report, Project No. S425). Washington, DC: U.S. Department of Health, Education, and Welfare.

Goodman, K. S. & Goodman, Y. M. (1979). Learning to read is natural. In L. B. Resnick & P. Weaver (Eds.), *Theory and practice of early reading*. Hilladale, NJ: Erlbaum.

Goodman, Y. M. (1967). *A psycholinguistic description of observed oral reading phenomena in selected young beginning readers*. Unpublished doctoral dissertation, Wayne State University, Detroit.

Goodman, Y. M. (1984). The development of initial literacy. In H. Goelman, A. Oberg, & F. Smith (Eds.), *Awakening to literacy*. Exeter, NH: Heinemann.

Gray, W. S. (1969). *The teaching of reading and writing* (2nd ed.). Glenview, IL: Unesco Scott, Foresman, & Co.

Hall, G. S. (1904). *The Psychology of Adolescence*. New York: D. Appleton.

Hildreth, G. H., Griffiths, N. L., & McGauvran, M. E. (1964). *Metropolitan readiness tests*. New York: Harcourt, Brace & World.

Hildreth, G. H., Griffiths, N. L., & McGauvran, M. E. (1966). *Metropolitan readiness tests*. New York: Harcourt, Brace & World.

Hildreth, G. H., Griffiths, N. L., & McGauvran, M. E. (1969). *Metropolitan readiness tests*. New York: Harcourt, Brace & World.

Holmes, M. C. (1927). Investigation of reading readiness of first grade entrants. *Childhood Education, 3*, 215-221.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. New York: The Macmillan Co.

Mason, J. M. (1984). Early reading from a developmental perspective. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 505-543). New York: Longman.

Meisels, S. J. (1987). Uses and abuses of developmental screening and school readiness testing. *Young Children, 42*(2)

Metropolitan Readiness Tests (1933). New York: World Book Co.

Mitchell, J. V., Jr. (Ed.) (1985). *The ninth mental measurements yearbook* (Vols. 1-2). Lincoln, NB: University of Nebraska-Lincoln, Buros Institute of Mental Measurements.

Monroe, M. (1932). *Children who cannot read*. Chicago: University of Chicago Press.

Morphett, M. V., & Washburne, C. (1931). When should children begin to read? *Elementary School Journal, 31*, 496-508.

Murphy, H. A. & Durrell, D. D. (1964). *Murphy-Durrell reading readiness analysis*. New York: Harcourt, Brace & World.

Plato (1892). The republic. In *The dialogues of Plato* (B. Jowett, Trans.) 2 vols. New York: Random House.

Shepard, L. A., & Smith, M. L. (1986). School readiness and kindergarten retention: A policy analysis. *Educational Leadership, 44*, 78-86.

Smith, N. B. (1928). Matching ability as a factor in first grade reading. *Journal of Educational Psychology, 19*, 560-571.

Teale, W. H., & Sulzby, E. (Eds.). (1986). *Emergent literacy: Writing and reading*. Norwood, NJ: Ablex.

Valencia, S. W., Pearson, P. D., & McGinley, W. (in press). Assessing reading and writing in the middle school classroom. In G. Duffy (Ed.) *Reading in the middle school* (2nd ed.). Newark, DE: International Reading Association.

Venezky, R.L. (1975). *Prereading skills: Theoretical foundations and practical applications*. (Theoretical Paper No. 54). Madison: The University of Wisconsin, Wisconsin Research and Development Center for Cognitive Learning.

Author Note

We gratefully acknowledge the contributions of Judith K. Shelton who produced the artwork for this report. The sample items in this report are similar to the items found on published tests in order to give the reader a sense of what the tests are like. However, the actual items found on the tests are not reproduced in this report in order to preserve their security.

A modified version of this report appears in: Morrow, L. M., & Smith, J. K. (Eds.) (1990). *Assessment for Instruction in Early Literacy.* Englewood Cliffs, NJ: Prentice Hall.

Table 1

Complete List of Test Categories, Subcategories, and Areas

I. General Test Characteristics

- A. Copyright
- B. Age
- C. Administration
  1. Group
  2. Individual
- D. Time
- E. Format
  1. Checklist
  2. Multiple Choice
- F. Number of Items
- G. Stimulus Size

II. Skills Tested

- A. World Knowledge
  1. Vocabulary/Picture Identification
  2. Sequencing
- B. Sound and Symbol Concepts
  1. Visual Matching and Embedded Patterns
  2. Upper Case Letters
  3. Lower Case Letters
  4. Matching Upper and Lower Case Letters
  5. Upper and Lower Case Letters--Mixed
  6. Letter - Sound Correspondences
  7. Initial Consonant Sounds
  8. Final Consonant Sounds
  9. Consonant Digraphs
  10. Consonant Blends
  11. Short Vowel Sounds
  12. Long Vowel Sounds
  13. Vowel Digraphs
  14. Dipthongs
  15. Phonetic Rules
  16. Syllables
  17. Spelling
  18. Rhyming
  19. Auditory Discrimination
  20. Memory
  21. Printing

Table 1 (Continued)

C. Literacy and Language Concepts
   1. School Language
   2. Syntax (including Negatives)
   3. Compound Words
   4. High Frequency Words
   5. Word Recognition (auditory)
   6. Decoding Printed Words in Spoken Context
   7. Environmental Print
   8. Orientation
   9. Directionality
   10. Concept of a Letter
   11. Concept of a Word
   12. Message
   13. Alphabetical Order

D. Comprehension
   1. Auditory Comprehension
   2. Reading

E. Others
   1. Colors
   2. Shapes
   3. Mathematics
   4. Quantitative Language

III. Presentation

A. Mode
   1. Auditory
   2. Visual
   3. Production

B. Unit of Presentation
   1. Graphemic
   2. Phonemic
   3. Syllabic
   4. Word
   5. Phrase
   6. Sentence
   7. Connected Discourse
   8. Patterns
   9. Numerical
   10. Picture
   11. Book

Table 1 (Continued)

IV. Response

A. Level of processing
   1. Recognition
   2. Identification
   3. Production

B. Unit of Response
   1. Picture
   2. Letter
   3. Word
   4. Nonsense Word
   5. Phrase
   6. Sentence
   7. Connected Discourse
   8. Book
   9. Numerical
   10. Objects

C. Student Response Activity
   1. Underline
   2. Mark
   3. Point
   4. Fill in the Circle
   5. Write
   6. Draw
   7. Manipulate
   8. Oral

Table 2

A Composite List of Skills/Behaviors/Traits Typically Found on Developmental Checklists

I. Physical Development

A. Large muscle development
1. Climbs on climbing equipment
2. Hops on one foot
3. Bounces a ball
4. Catches a ball
5. Throws a ball
6. Walks on tip toes
7. Walks on a balance beam
8. Jumps from a stool
9. Skips
10. Gallops
11. Claps in time to music

B. Fine muscle development
1. Fastens buttons
2. Ties shoelaces
3. Strings beads
4. Puts puzzles together
5. Colors in the lines
6. Cuts with scissors
7. Copies letters and shapes

C. Sensory development
1. Recognizes objects drawn on board from across room
2. Holds book at appropriate distance for reading
3. Follows pictures in sequence across page with eyes
4. Responds to oral directions from across the room

D. General health
1. Appears to be well nourished
2. Remains awake and alert throughout the day
3. Enters into outdoor play energetically
4. Attendance records indicate adequate health

II. Language development

A. Speech development
1. Speaks clearly
2. Expresses needs adequately
3. Speaks language of school fluently
4. Talks about everyday experiences
5. Answers willingly when spoken to
6. Modulates voice
7. Asks questions

Table 2 (Continued)

B. Vocabulary development
   1. Asks about meanings of unfamiliar words
   2. Labels objects and actions correctly
   3. Describes objects in terms of size, shape, color
   4. Uses relationship words appropriately
   5. Uses appropriate category labels

C. Language development
   1. Understands language of school
   2. Follows three part oral directions correctly
   3. Completes sentences with a logical ending
   4. Completes story with a logical ending
   5. Retells a familiar story
   6. Discriminates between words

D. Written language
   1. Shows an interest in reading
   2. Attempts to read
   3. Asks about words written around the room
   4. Recognizes own name when written
   5. Attempts to write own name
   6. Dictates a phrase or sentence about an experience
   7. Knows reading progression
      a. Left to right
      b. Top to bottom

III. Cognitive development

A. Information
   1. Names colors
   2. Names shapes
   3. Names sizes
   4. Knows the five senses
   5. Names body parts
   6. Gives own name
   7. Gives own address
   8. Gives own telephone number

B. Awareness of details
   1. Recognizes likenesses and differences of objects
   2. Recognizes likenesses and differences of people
   3. Includes details when drawing a person

C. Memory
   1. Memorizes names of other children
   2. Memorizes songs and rhymes
   3. Recites the alphabet
   4. Counts to twenty

Table 2 (Continued)

D. Temporal and spatial concepts
   1. Names days of the week
   2. Names seasons of the year
   3. Uses today, tomorrow, yesterday correctly
   4. Tells time to the hour
   5. Tells time to the half hour
   6. Identifies right and left on own body
   7. Identifies right and left in space

IV. Social-emotional development

A. Intrapersonal skills
   1. Recognizes own property
   2. Keeps track of own property
   3. Tries new activities
   4. Tolerates a reasonable amount of frustration
   5. Shows pride
   6. Demonstrates responsibility
   7. Shows creativity
   8. Listens to a story for at least 15 minutes
   9. Listens to directions before responding
   10. Perseveres for at least 10 min. on a single task
   11. Completes tasks
   12. Works alone without distraction
   13. Responds positively to change in routine
   14. Find way from school to bus, carpool, or home

B. Interpersonal skills
   1. Plays cooperatively/competitively as appropriate
   2. Takes turns
   3. Helps children and adults spontaneously
   4. Follows adult direction without complaint
   5. Leaves parents with little or no reluctance
   6. Seeks help when needed
   7. Responds appropriately to the emotions of others
   8. Shows feelings
   9. Protects self
   10. Takes lead in playing with younger children
   11. Participates in conversations

C. Hygiene/self help
   1. Toilet trained
   2. Dresses self
   3. Tries new food

Table 3

Average Number of Items Per Subtest (and Number of Subtests) by Test and Major Skill Category

Test	World Knowledge	Sounds and Symbols	Literacy and Language Concepts	Comprehension	Other Skills
CTBS A	8.0 (1)	20.50 (6)	14.00 (2)	2.00 (1)	26.00 (1)
CTBS B		5.75 (4)	6.25 (4)	16.00 (2)	20.00 (1)
Metropolitan 2		11.20 (5)	9.00 (1)	9.00 (1)	12.00 (2)
Metropolitan 1		8.67 (6)	15.00 (1)		11.00 (1)
Roswell/Chall		9.71 (14)	35.00 (1)		
MAT 6		13.50 (4)	24.00 (1)		24.00 (1)
SESAT 1	27.50 (2)	6.00 (8)		37.50 (2)	42.00 (1)
SESAT 2	27.50 (2)	8.00 (5)		25.00 (4)	44.00 (1)
Lollipop		4.33 (3)	9.00 (1)	1.00 (1)	4.14 (7)
Boehm R			50.00 (1)		
PRI a	5.33 (3)	2.67 (3)	4.00 (5)	8.00 (1)	
ConcAbPrnt		1.00 (1)	2.22 (9)		
ReadyToRead			45.00 (1)		
GRTR Survey			13.00 (2)		
GRTR Diag	9.00 (1)	16.75 (4)	9.00 (1)		
Circus Batt	35.50 (2)	8.67 (15)	26.67 (3)	15.00 (4)	20.00 (3)
CAT 11		8.50 (4)	14.00 (4)	11.67 (3)	54.00 (1)
CAT 12	10.00 (2)	9.20 (5)	10.00 (2)	11.00 (2)	20.00 (1)
ITBS 6	29.00 (1)	4.56 (9)	22.67 (3)	16.00 (4)	33.00 (1)
ITBS 5	29.00 (1)	7.00 (5)	29.00 (1)	31.00 (1)	33.00 (1)
Totals					
Number of subtests	15	100	44	26	22
Mean Number of items	19.46	8.95	13.39	16.88	19.09
(St.Dev)	13.42	6.17	12.46	12.23	15.86
Percent of Total Items	11.03	34.27	22.25	16.58	15.87

Table 4

Frequency of Skills Tested and Percentage of Total

	Frequency	Percent
A. World Knowledge		
1. Vocabulary/Picture Identification	14	6.7
2. Sequencing	1	0.5
B. Sound and Symbol Concepts		
1. Visual Matching	9	4.3
a. Embedded Patterns	1	0.5
2. Upper Case Letters	11	5.3
3. Lower Case Letters	9	4.3
4. Matching Upper and Lower Case Letters	3	1.4
5. Upper and Lower Case Letters - Mixed	3	1.4
6. Letter-Sound Correspondences	10	4.8
7. Initial Consonant Sounds	18	8.7
8. Final Consonant Sounds	7	3.4
9. Consonant Digraphs	1	0.5
10. Consonant Blends	2	1.0
11. Short Vowel Sounds	6	2.9
12. Long Vowel Sounds	3	1.4
13. Vowel Digraphs	2	1.0
14. Dipthongs	1	0.5
15. Phonetic Rules	1	0.5
16. Syllables	1	0.5
17. Spelling	1	0.5
18. Rhyming	5	2.4
19. Auditory Discrimination	2	1.0
20. Memory	2	1.0
21. Printing	2	1.0
C. Literacy and Language Concepts		
1. School Language	11	5.3
2. Syntax	9	4.3
a. Negatives	1	0.5
3. Compound Words	1	0.5
4. High Frequency Words	11	5.3
5. Word Recognition (auditory)	2	1.0
6. Decoding Printed Words in Spoken Context	1	0.5
7. Orientation	1	0.5
9. Directionality	2	1.0
10. Concept of a Letter	1	0.5
11. Concept of a Word	1	0.5
12. Message	1	0.5
13. Alphabetical Order	1	0.5
D. Comprehension		
1. Auditory Comprehension	19	9.1
2. Reading	7	3.4

Table 4 (Continued)

	Frequency	Percent
E. Others		
1. Colors	2	1.0
2. Shapes	4	1.9
3. Mathematics	14	6.7
4. Quantitative Language	2	1.0

Table 5

Average Number of Items (and Number of Subtests) by Type of Administration, Major Skill Area, and Level of Processing Required of Students

	Major Skills Area					
Level of Processing	World Knowledge	Sounds & Symbols	Literacy & Language	Comprehension	Other Skills	Total
			Group Tests			
Recognition	19 (16)	9.71 (72)	14.48 (23)	22.8 (15)	25.58 (12)	
Identification	-----	5.5 (8)	13.29 (7)	11.75 (8)	34.5 (2)	
Production	-----	1.0 (2)	27.0 (2)	1.0 (2)	15.0 (1)	
			Individual Tests			
Recognition	-----	5.0 (1)	2.5 (6)	-----	3.0 (3)	
Identification	-----	8.7 (14)	15.67 (6)	1.0 (1)	5.67 (3)	
Production	-----	7.33 (3)	-----	-----	3.0 (1)	
			Total (Group + Individual)			
Recognition	19 (16)	9.64 (73)	12.0 (29)	22.8 (15)	21.07 (15)	13.61 (148)
Identification	-----	7.59 (22)	14.38 (13)	10.55 (9)	17.2 (5)	10.92 (49)
Production	-----	4.8 (5)	27.0 (2)	1.00 (2)	9.0 (2)	8.91 (11)
Total	19 (16)	8.95 (100)	13.39 (44)	16.88 (26)	19.09 (22)	

Table 6

Frequency of Subtests by Units of Presentation and Units of Response

Unit of Response	Unit of Presentation											
	A	B	C	D	E	F	G	H	I	J	Sum	%
picture	---	26	21	---	26	12	---	3	1	5	94	45.9
letter	28	11	---	---	---	---	---	---	---	---	39	19.0
word	1	8	21	---	3	---	---	---	---	---	33	16.1
phrase	---	---	---	1	---	---	---	---	---	---	1	0.5
sentence	---	---	1	---	7	1	---	---	---	---	9	4.4
connected discourse	2	---	1	---	---	4	---	---	---	---	7	3.4
book	---	---	---	---	---	---	4	---	---	---	4	2.0
numerical	---	---	---	---	---	---	---	---	---	8	8	3.9
patterns/ shapes	1	---	---	---	---	---	---	---	8	---	9	4.4
objects	---	---	---	---	1	---	---	---	---	---	1	0.5
not classified	1	---	1	---	---	---	---	1	---	---	3	
Total	33	45	45	1	37	17	4	4	9	13	208	
% of Total	15.9	21.6	21.6	0.5	17.8	8.2	1.9	1.9	4.3	6.3		

A graphemic
B phonemic
C word
D phrase
E sentence
F connected discourse
G book
H picture
I patterns/shapes
J numerical

Table 7

Frequency of Subtests Using Different Response Activities

Response Activity	Frequency	Percent
fill in circle	131	63.0
mark	24	11.5
oral	24	11.5
point	12	5.8
underline	8	3.8
write	7	3.4
manipulate	1	.5
draw	1	.5
Total	208	100

Table 8

Frequency of Subtests Using Different Modes of Presentation

Mode of Presentation	Frequency	Percent
auditory	154	74.0
visual	52	25.0
writing	2	1.0
Total	208	100

Figure 1
World Knowledge Items

Example 1a
Item testing world knowledge

What the child sees:

What the child hears: *Look at the pictures of the three trucks. Listen Carefully. Yesterday there was a fire at Jones' Garage. Mrs. Jones called the fire department. Which truck came to put out the fire?*

Example 1b
Item testing vocabulary knowledge

What the child sees:

What the child hears: *There are pictures of a toaster, an iron, and a fan. Mark the oval under the picture of the fan.*

Figure 2
Items Testing Knowledge of Sound and Symbol Concepts

1. Visual Matching

What the child sees: **b | f d b g**

What the child hears: *Find the letter that is exactly the same as the first letter.*

2. Auditory Discrimination of Words with Similar Parts

What the child sees:

What the child hears: *The pictures are book, drill, and pencil. Find the picture that begins with the same sound that you hear at the beginning of bottle.*

3. Letter-Sound Correspondence

What the child sees:

i
b
q

What the child hears: *Find the letter that stands for the sound you hear at the beginning of bus.*

4. Letter recognition

What the child sees:

p b q o

What the child hears: *Find the b.*

5. Letter identification

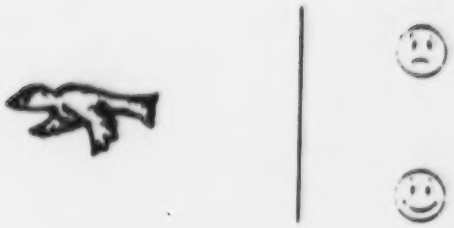
What the child sees: **z x c v b n m**
What the child hears: *Name these letters.*

6. Letter Production

What the child sees: Nothing
What the child hears: *Write the letter b.*

Figure 3
Item Testing Language Usage

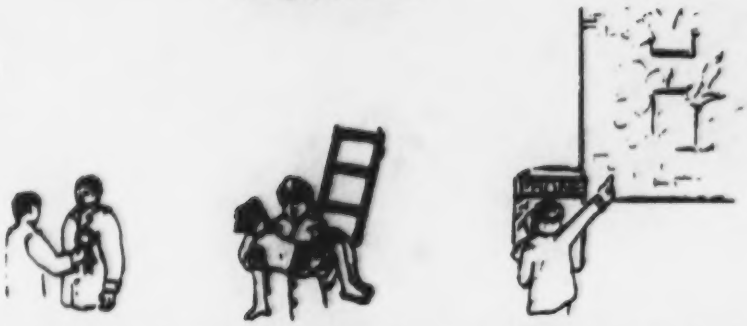
What the child sees:



What the child hears: *Listen to what I say. The bird are flying. Mark the happy face if I said it right. Mark the sad face if I said it wrong.*

Figure 4
Comprehension

Example a.
What the child sees:



What the child hears: *Everyone on the block was standing outside the apartment building when the fire trucks came. All of the people were out of the burning building except a woman on the third floor. The firefighters were able to bring the woman down safely. Everyone cheered when they got to the ground. Mark the picture that shows what the people cheered about.*

Example b.
What the child sees:

What the child hears: *The three pictures show some roller skates, an umbrella, and a snowsuit. When the Jones children cleaned out their drawers and closets for the garage sale, they found many things they didn't need any more. Sarah sold some stuffed toys and a snowsuit. Lisa sold a pair of roller skates with a broken strap and some pants that no longer fit. Mike sold an old umbrella and some comic books. Mark the picture of something Lisa sold.*

Example c.
What the child sees:

The cat came _____ .

home can house boat

What the child hears: *Read the sentence and the four words under the sentence. Mark the word that goes best in the blank in the sentence.*

Figure 5
Other Types of Items

Example 1: Mathematics

What the child sees:

What the child hears: ***Mark the box that has four apples in it.***

Example 2: Quantitative Language

What the child sees:

What the child hears: ***Look at the three butterflies. Mark the picture of the largest butterfly.***

Figure 6
Units of Presentation and Units of Response

A. Presentation: Graphemic **Response: Graphemic**

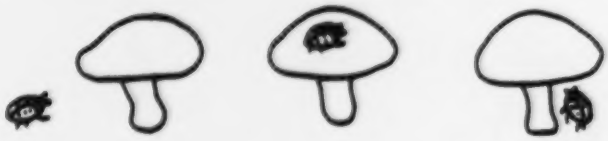
What the child hears: ***Mark the letter that has the same name as the first letter.***

B. Presentation: Word **Response: Picture**

"Candle"

What the child hears: ***Candle.***

C. Presentation: Sentence **Response: Picture**



What the child hears: ***Mark the picture of the bug that is sitting under the mushroom.***

D. Presentation: Pattern **Response: Pattern**

What the child hears: ***Mark the pattern that is exactly the same as the first pattern.***

Figure 7
An Item Patterned After the
Metropolitan Readiness Test: Memory Subtest

What the child sees:

What the child hears: *(Find the box that has...) Chair, house, tree.*

Figure 8
An Item Patterned After the
Metropolitan Readiness Test: Auditory Discrimination of Beginning Consonant Sounds

What the child sees:

What the child hears: *The pictures are, cat, sock, milk, and arm. Find the one that begins with the same sound as miss and mine.*

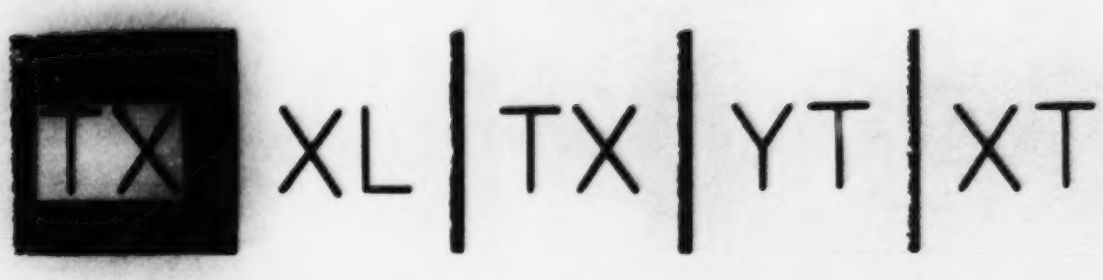
Figure 9
An Item Patterned After the
Metropolitan Readiness Test: Letter Recognition

What the child sees:

What the child hears: *Mark the m.*

Figure 10
An Item Patterned After the
Metropolitan Readiness Test: Visual Matching

What the child sees:



What the child hears: *Nothing, but the general direction is to mark under the one that is just like what you see in the first* (actually they give the color) *box.*

Figure 11
An Item Patterned After the
Metropolitan Readiness Test: School Language

What the child sees:

What the child hears: *Mark the picture that shows the dog between the children.*

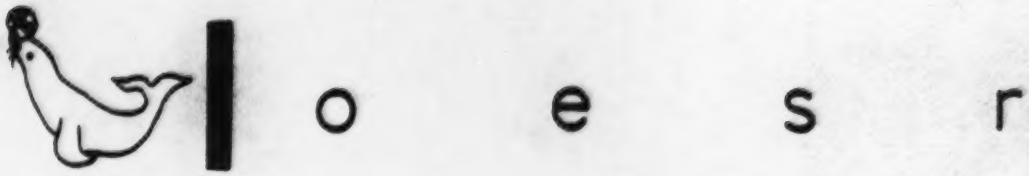
Figure 12
An Item Patterned After the
Metropolitan Readiness Test: Quantitative

What the child sees:

What the child hears: *Mark the second duck.*

Figure 13
An Item Patterned After the
Metropolitan Readiness Test: Letter-Sound Correspondences

What the child sees:



What the child hears: *Listen to the sound at the beginning of seal. Mark under that letter.*

Figure 14
An Item Patterned After
Basal Readiness Tests

What the child sees:

What the child hears: *Listen to the sound at the beginning of seal. Mark that letter.*